

How search engines disseminate information about COVID-19 and why they should do better

Makhortykh, Mykola; Urman, Aleksandra; Ulloa, Roberto

Veröffentlichungsversion / Published Version

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Makhortykh, M., Urman, A., & Ulloa, R. (2020). How search engines disseminate information about COVID-19 and why they should do better. *Harvard Kennedy School Misinformation Review*, 1(3). <https://doi.org/10.37016/mr-2020-017>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>



Research Article

How search engines disseminate information about COVID-19 and why they should do better

Access to accurate and up-to-date information is essential for individual and collective decision making, especially at times of emergency. On February 26, 2020, two weeks before the World Health Organization (WHO) officially declared the COVID-19's emergency a "pandemic," we systematically collected and analyzed search results for the term "coronavirus" in three languages from six search engines. We found that different search engines prioritize specific categories of information sources, such as government-related websites or alternative media. We also observed that source ranking within the same search engine is subjected to randomization, which can result in unequal access to information among users.

Authors: Mykola Makhortykh (1), Aleksandra Urman (2), Roberto Ulloa (3)

Affiliations: (1, 2) Institute of Communication and Media Studies, University of Bern, (3) GESIS – Leibniz Institute for the Social Sciences

How to cite: Makhortykh, M.; Urman, A.; Ulloa, R. (2020). How search engines disseminate information about COVID-19 and why they should do better, *The Harvard Kennedy School (HKS) Misinformation Review*, Volume 1, Special Issue on COVID-19 and Misinformation

Received: March 26th, 2020 Accepted: May 1st, 2020 Published: May 11th, 2020

Research questions

- How do search engines select and prioritize information related to COVID-19?
- What is the impact and consequences of the randomization on information ranking and filtering mechanisms?
- How much do the above-mentioned aspects of web search vary depending on the language of the query?

Essay summary

- Using multiple (N=200) virtual agents (i.e., software programs), we examined how information about the coronavirus is disseminated on six search engines: Baidu, Bing, DuckDuckGo, Google, Yandex, and Yahoo. We scripted a sequence of browsing actions for each agent and then tracked these actions under controlled conditions, including time, agent location, and browser history.
- On February 26, 2020, the agents simultaneously entered search queries based on the most common term used to refer to the COVID-19 pandemic (i.e., "coronavirus" in English, Russian,

¹ A publication of the Shorenstein Center for Media, Politics, and Public Policy, at Harvard University, John F. Kennedy School of Government.

and Mandarin Chinese) into the six search engines.

- The analysis of the search results acquired by the agents highlighted unsettling differences in the types of information sources prioritized by different engines. We also identified a considerable effect of randomization on how sources are ranked within the same search engine.
- Such discrepancies in search results can misinform the public and limit the rights of citizens to make decisions based on reliable and consistent information, which is of particular concern during an emergency, such as the COVID-19 pandemic.

Implications

We identified large discrepancies in how different search engines disseminate information about the COVID-19 pandemic. Some differences in the results are expected given that search engines personalize their services (Hannak et al., 2013), but our study highlights that even non-personalized search results differ substantially. For example, we found that some search algorithms potentially prioritize misleading sources of information, such as alternative media and social media content in the case of Yandex, while others prioritize authoritative sources (e.g., government-related pages), such as in the case of Google.

The randomization of search results among users of the same search engine is of particular concern. We found that the degree of randomization varies between the engines: for some, such as Google and Bing, it mostly affects the composition of the “long tail” of search results, such as those below the top 10 results, while others, such as DuckDuckGo and Yandex, also randomize the top 10 results. Randomization ensures that what a user sees is not necessarily what the user chooses to see, and that different users are exposed to different information. Through randomization, a user sees what the search engine randomly decided that that specific user is allowed to see. Then, in this scenario, access to reliable information is simply a matter of luck.

While randomization can encourage knowledge discovery by diversifying information acquired by individuals (Helberger, 2011), it can be detrimental when the society urgently needs to access consistent and accurate information – such as during a public health crisis. If we assume that a major driver of randomization is the maximization of user engagement by testing different ways of ranking search results and choosing the optimal hierarchy of information resources on a specific topic (e.g., the so-called “Google Dance” (Battelle, 2005)), then we would be in a situation in which companies’ private interests directly interfere with the people’s rights to access accurate and verifiable information.

The exact functioning of – and justification for – randomization and different source priorities is currently unknown. Criticism of algorithmic non-transparency in information distribution is not new (Pasquale, 2015; Kemper & Kolkman, 2019; Noble, 2018). However, lack of transparency is particularly troublesome in times of emergency when the biases of filtering and ranking mechanisms become a matter of public health and national security. Our observations show that search engines retrieve inconsistent and sometime misleading results in relation to COVID-19, but it remains unclear what factors contribute to these information discrepancies and what principles each engine uses to construct hierarchies of knowledge. These issues raise multiple questions, including what is “good” information, who should decide on its quality and can these decisions be applied univocally. Most importantly: should search engines suspend randomization in times of public emergencies?

Finding answers to these questions is not easy and will require time and appropriate efforts. One starting point to improve search transparency could be to make resources for conducting “algorithmic auditing” (that is, analyses of algorithmic performance similar to the one implemented in this study) more accessible to the academic community, and the public at large (Mittelstadt, 2016). Currently, there is no openly available and scalable infrastructure that can be used to compare the performance of different search engine algorithms, as well as their particular features (e.g., randomization). By providing such

infrastructure, and making data on the effects of algorithms on information distribution more accessible, search engine companies could address the lack of transparency in their algorithmic systems and increase trust in that information technologies play in our societies (Foroohar, 2019).

Another point to consider is the possibility of implementing “user control mechanisms” that can ensure that search engine users can tackle algorithmic features (e.g., randomization) interfering with their ability to receive information (He, Parra, & Verbert, 2016; Harambam et al., 2019). User-centric approaches can vary from clear policies towards source prioritization (e.g., Google’s decision to prioritize government-related sources on COVID-19, but applied consistently to other search subjects) to an option to opt out not only from search personalization, but also its randomization.

Findings

Finding 1: Your search engine determines what you see

We found large discrepancies in the search results ($N \approx 50$) between identical agents using different search engines (Figure 1). Despite the use of the same search queries, all the metrics showed less than 25% similarity in search results between the engines, except DuckDuckGo and Yahoo, which shared almost 50% of their results. In many cases, we observed almost no overlap in the search results (e.g., between Google and DuckDuckGo), thus indicating that users receive completely different selections of information sources. While differences in source selection are not necessarily a negative aspect, the complete lack of common resources between the search engines can result in substantial information discrepancies among their users, which is troubling during an emergency. Furthermore, as Finding 3 shows, search engines prioritize not just different sources of the same type (e.g., various legacy media outlets) but different types of source, which has direct implications for the quality of information that the engines provide.

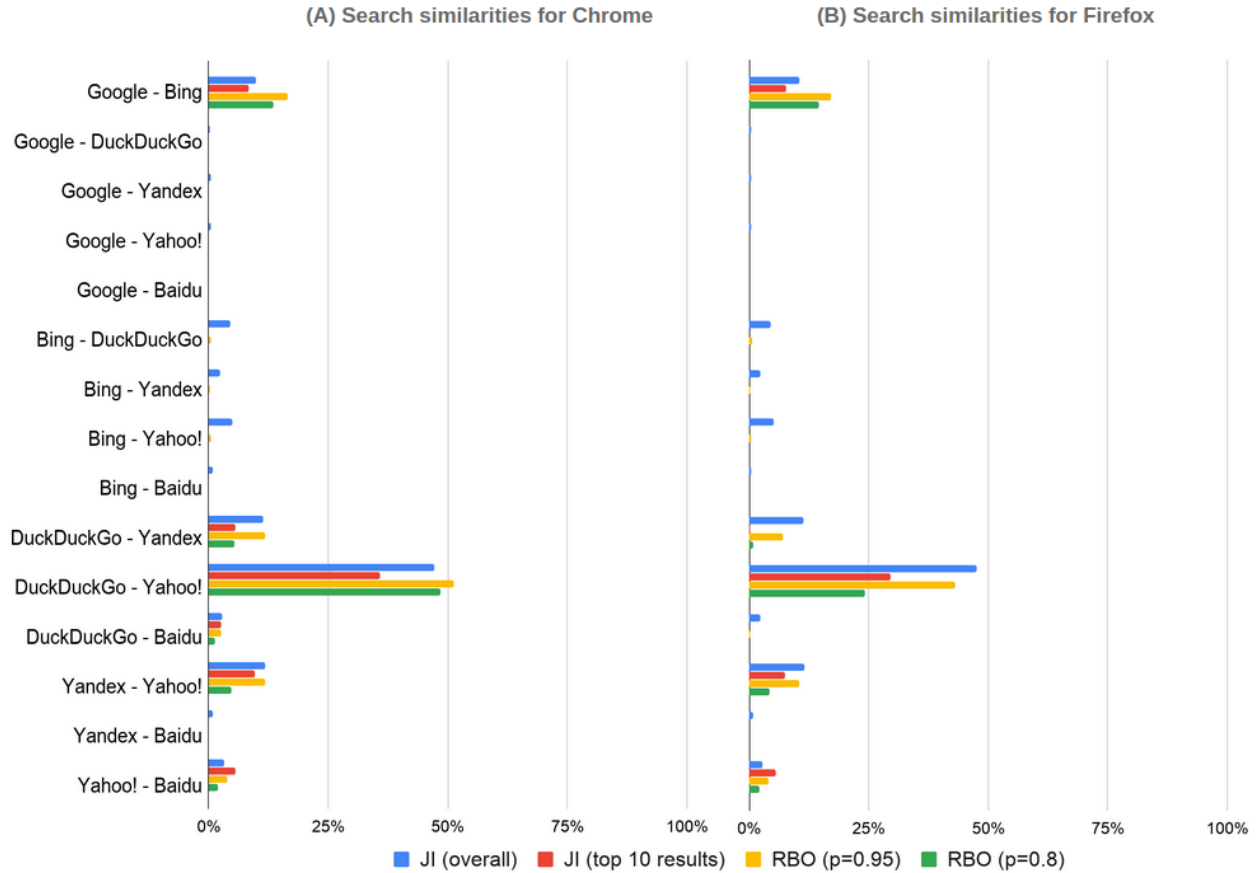


Figure 1. Average similarities (%) in search results for pairs of agents using different search engines. *JI* represents the Jaccard index and *RBO* stands for the Ranked Biased Overlap. The x-axis shows the average percentage for each of the similarity metrics, and the y-axis shows the pairs of search engines that are compared.

The Jaccard index (*JI*), a metric that measures the share of common results between different agents, showed that for most engines, the source overlap occurred in the long tail of results, that is, those beyond the top 10 results. In the top 10 results, the overlap was higher only for the Yahoo-Baidu pair; the rest of the results comprised largely different sources. These observations were supported by the Ranked Biased Overlap (*RBO*), a metric that considers the ranking of search results. The parameter p determines how important the top results are: $p = 0.8$ gives more weight to the few top results, whereas $p = 0.95$ distributes the weight more equally between the top ~ 30 results. Our *RBO* values suggested that the ranking of the long-tail results was usually more consistent between the search engines than the ranking of the top search results.

Finding 2: The search results you receive are randomized

We observed substantial differences in the search results received by the identical agents using the same search engine and browser (Figure 2). For some search engines, such as Yahoo and Baidu, we found substantial consistency in the composition of the general and top 10 results (as indicated by the *JI* values). However, as indicated by their *RBO* values, the ranking of these results was inconsistent. By contrast, on Google and, to a certain degree, Bing, the top 10 results were consistent, whereas the rest of the results were less congruent. Finally, in the case of agents using DuckDuckGo in Chrome, the overall selection of the results was consistent, but their rankings varied substantially.

One possible explanation for such randomization is that the search algorithms introduce a certain

degree of “serendipity” into the way the results are selected to give different sources an opportunity to be seen (Cornett, 2017). A more pragmatic reason for randomization might be that the search engines are constantly experimenting with the results to determine which maximize user satisfaction and potentially increase profits through search advertisements. Such experimentation seems to be particularly intense in the case of unfolding and rapidly changing topics, for which there are no pre-existing knowledge hierarchies.

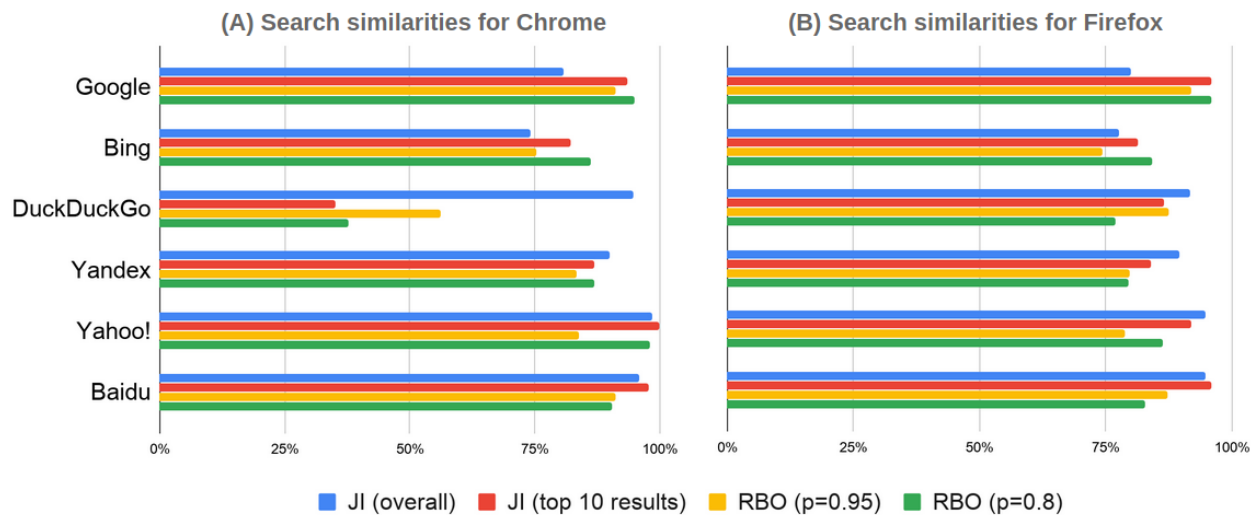


Figure 2. Average similarities (%) in search results for pairs of agents using the same search engine. JI stands for the Jaccard index and RBO stands for the Ranked Biased Overlap. The x-axis shows the average percentage for each of the similarity metrics, and the y-axis shows the pairs of search engines that are compared.

To test the later assumption, we ran a series of queries that were not related to the COVID-19 pandemic but to more established news topics. When we compared the results for “coronavirus” with other searches performed by bots (e.g. “US elections”; see Appendix), we observed higher volatility in the coronavirus results, which may be due to its novelty and the absence of historical information about user preferences.

Finally, we considered whether the type of browser influences the degree of randomization. For most search engines (except DuckDuckGo, where the ranking of results was strongly randomized in Chrome), we did not observe major differences between the browsers. These observations may indicate that the choice of browser does not have a substantial influence on the selection of results. At the same time, the lack of such influence can also be attributed to the recency of the coronavirus case, which translates to a lack of historical data based on which algorithms offer a more browser-specific selection of results.

Finding 3. Search engines prioritize different types of sources

We found that the search engines assign different priorities to specific categories of information sources in relation to the coronavirus. Our examination of the 20 most frequently occurring search results in English (Figure 3) indicated that most engines prioritized sources associated with legacy media (e.g., CNN) and healthcare organizations (including public health ones, such as WHO). However, the ratio of these sources varies substantially: for Bing, for instance, healthcare-related sources constituted almost half of the top 20 results, whereas for Google, Yandex, and Yahoo, they comprised less than one-quarter of the top results. By contrast, Yahoo prioritized recent information updates on the coronavirus from legacy media, while Google gave preference to government-related websites, such as those of city councils.

Considering their ability to spread unverified information, these differences in the knowledge

hierarchies constructed by the search engines are troubling. For some of the search engines, the top search results about the pandemic included social media (e.g., Reddit) or infotainment (e.g., HowStuffWorks). Such sources are generally less reliable than official information outlets or quality media. Moreover, in the case of Yandex, the top search results included alternative media (e.g., <https://coronavirusleaks.com/>), in which the reliability of information is questionable.

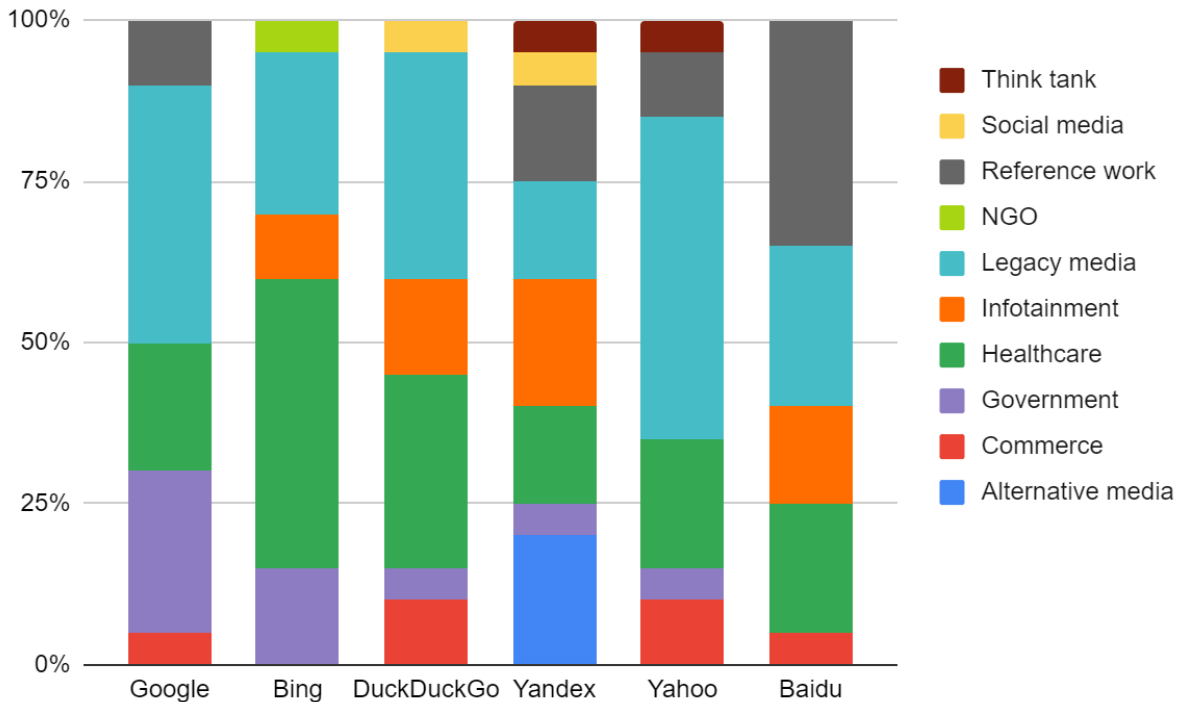


Figure 3. Top 20 results most frequently provided by each engine by category. The y-axis shows the ratio of different categories of sources for each engine.

Considering their ability to spread unverified information, these differences in the knowledge hierarchies constructed by the search engines are troubling. For some of the search engines, the top search results about the pandemic included social media (e.g., Reddit) or infotainment (e.g., HowStuffWorks). Such sources are generally less reliable than official information outlets or quality media. Moreover, in the case of Yandex, the top search results included alternative media (e.g., <https://coronavirusleaks.com/>), in which the reliability of information is questionable.

Finding 4. The choice of language affects differences between engines

Finally, we found that the degree of discrepancies between the search results varied depending on the language in which the queries were executed. The variation was relatively low among the agents using the same search engine, suggesting that the choice of language did not have a substantial influence on randomization (except for Google, where results in Chinese/Russian were more randomized than those in English).

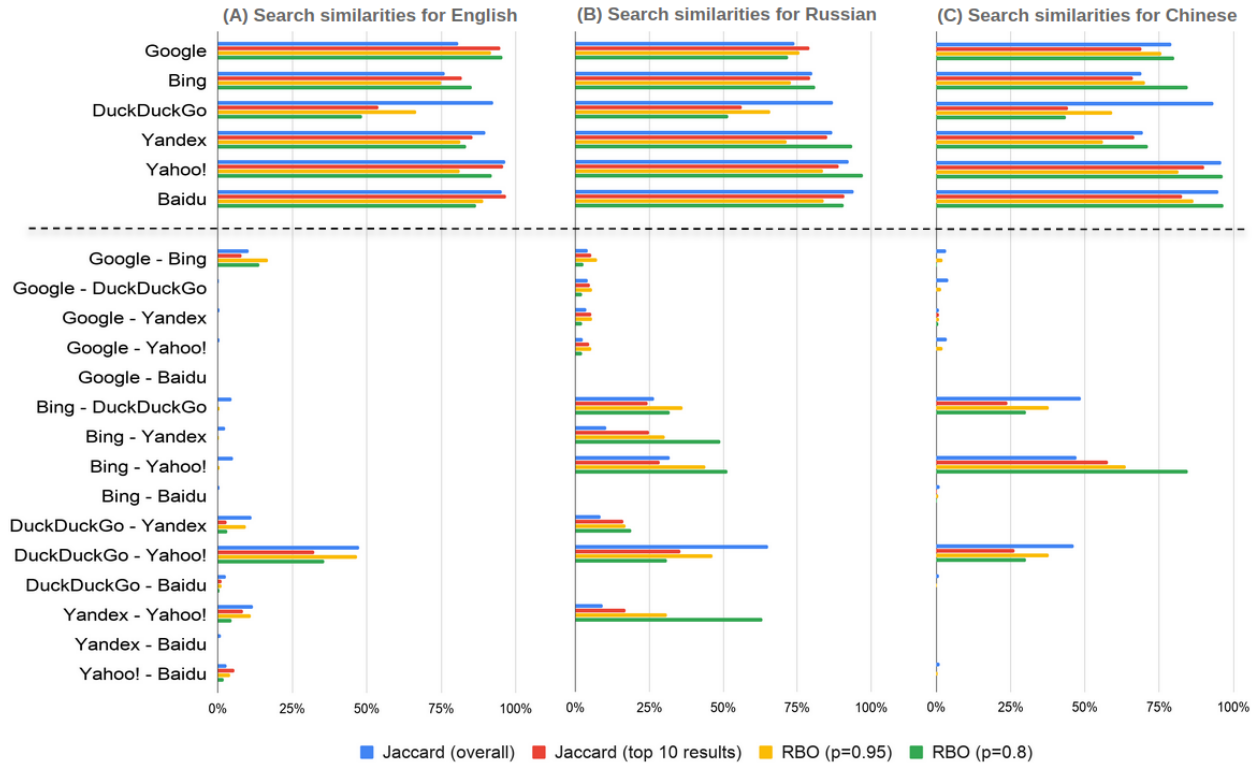


Figure 4. Average similarities (%) for paired agents according to search engines across languages. *Jl* stands for the Jaccard index and *RBO* stands for the Ranked Biased Overlap. The x-axis shows the average percentage for each of the similarity metrics, and the y-axis shows the pairs of search engines that are compared.

We observed that language choice led to more substantial variation between the pairs of search engines (Figure 4). For several pairs, such as Bing and DuckDuckGo, we found that search results were more similar for Russian and Chinese than for English. This pattern, however, was not universal, as shown by the results for the “US elections” query (Figure A1 in the Appendix). These observations may be explained by different volumes of information in the respective languages processed by the search engines. The smaller volume of data in Russian and Chinese may increase the probability of sources overlapping. Nonetheless, these language-based discrepancies support our general claim of the lack of consistency among search engines.

Methods

Research questions. In our study, we looked at three research questions: 1) How do search engines select and prioritize information related to COVID-19? 2) To what degree are the ranking and filtering mechanisms within the same search engine influenced by randomization? 3) How much do the above-mentioned aspects of web search vary depending on the language of the query?

Data collection. We developed a distributed cloud-based research infrastructure to emulate and track the browsing behavior of multiple virtual agents. Using Amazon Elastic Compute Cloud (Amazon EC2), we created 100 CentOS virtual machines with two browsers (Firefox and Chrome) installed on each machine. In each browser (the “agent”), we installed two browser extensions: a bot and a tracker. The bot emulates browsing behavior by searching terms from a pre-defined list and then navigates through the search results. The tracker collects the HTML from each page visited by the bot and sends it to a central server, a different machine in which all data are stored.

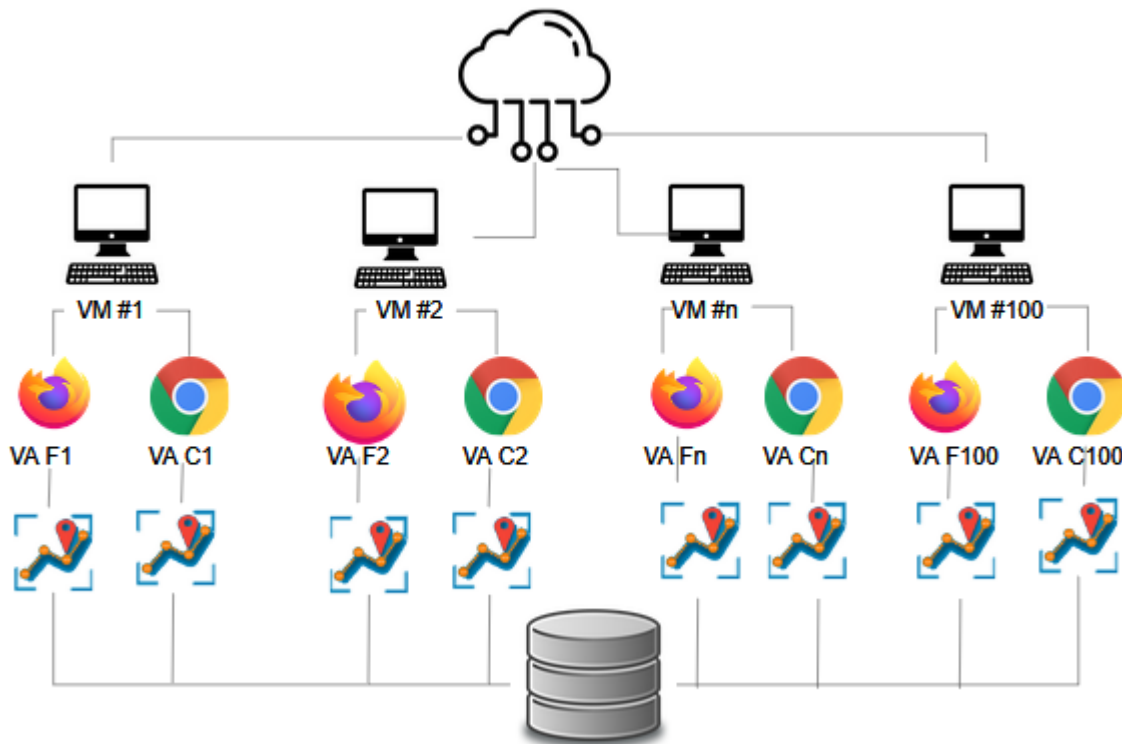


Figure 5. Data collection infrastructure schema

We looked at six search engines—Baidu, Bing, Google, Yahoo, Yandex and DuckDuckGo—and queried them for the term “coronavirus” in English, Russian, and Chinese. The choice of the first five search engines is attributed to them being the most commonly used worldwide (see Statista 2020), whereas DuckDuckGo was chosen because of its focus on privacy (Schofield 2019) that translates into it not using users’ personal data to personalize search results (Weinberg, 2012). We chose the word “coronavirus” as the most common term used in relation to the ongoing pandemic at the time of data collection (see Google Trends (2020)) and the one denoting a broad family of viruses that includes COVID-19. In all cases, we used the fixed spelling of the term (i.e., “coronavirus” and not “corona virus”) to prevent query discrepancies from influencing the search results.

We divided all agents into six groups ($n=32/33$; equally distributed between Chrome and Firefox), each of which was assigned to a specific search engine. On February 26, 2020, two weeks before WHO declared the coronavirus a pandemic, each agent performed a similar browsing routine consisting of typing, clicking, and scrolling actions typical of web search navigation. The length of the routine was kept under three minutes to enable the collection of a comparable amount of data (~50 top search results). All agents were synchronized to start each query at exactly the same time (i.e., every seven minutes). In the case of a network failure, a page refresh was triggered to enable re-synchronization of the agent for the next round of queries.

We also cleaned browser data at the beginning of each round of queries to reduce the influence of previous searches on the subsequent ones. First, we removed data that can be accessed by the search engine JavaScript, comprising local storage, session storage, and cookies. Second, we removed data that can only be accessed by the browser or extensions with advanced privileges (see Table A1 in the Appendix for a full list), which included browser history and cache.

Data analysis. After collecting the data, we used BeautifulSoup (Python) and rvest (R) to extract search result URLs from HTML for a given query and to filter nonrelevant URLs (e.g., digital ads). While we acknowledge that nonrelevant URLs can also influence the way that users perceive information about the

coronavirus, we decided to remove them because of our interest in the default mechanisms for search filtering and ranking.

We then conducted search results between each pair of two agents ($200 \times 199 / 2 = 19900$ pairs) using two similarity metrics: JI and RBO. JI measures the overlap between two sets of results and can be defined as a share of the identical results in both sets among all the unique results. It is frequently used to audit search algorithms (Hannak et al., 2013; Kliman-Silver et al., 2015; Puschmann, 2019). The value of JI varies between 0 and 1, with 1 indicating that the compared sets are 100% identical and 0 that they are completely different. We calculated two JIs for each pair of agents: one measuring the overlap between the full sets of search results from each agent and one measuring the overlap for the top 10 results.

JI usefully assesses how many URLs are repeated within the search results acquired by the agents, but it does not account for the order of these URLs. However, source ranking is crucial for auditing search algorithms: the composition of two sets of results can be identical (hence, $JI = 1$), but the order of these results can be different, leading to incompatible user experiences. Discrepancies in ranking are particularly important because users rarely scroll below the few top results (Cutrell & Guan, 2007; González-Caro & Marcos, 2011).

To account for differences in search result ranking, we used RBO, a metric designed explicitly for investigating similarities in the output of search engines (Webber et al., 2010; Cardoso & Magalhães, 2011; Robertson et al., 2018). Unlike JI, RBO weights top search results more than lower ones to take into account that a source's ranking influences the probability of the user seeing it (Cutrell & Guan, 2007). Similar to JI, an RBO value of 1 indicates that the composition of results and their rankings are 100% identical, whereas an RBO value of 0 indicates that both the results and their rankings are completely different.

For RBO, the exact weight of the results' positioning is determined by the parameter p . The value of the parameter is decided by the researcher and can be any number within a $0 < p < 1$ range. The lower the value of p , the more weight is placed on the top results. We ran our analysis with two different values of p : 0.95 and 0.8. The former value enables a more systemic analysis of the ranking differences between the two agents, whereas the latter focuses on the variation between the few top results (Webber et al., 2010, p. 24).

After calculating similarity metrics, we examined the types of information source prioritized by the search engines. Using an inductive coding approach, we examined the 20 most common URLs among search results in English for each of the six engines and coded each URL according to the category of information source to which it belonged. We identified 10 types of information source: 1) *alternative media*: non-mainstream and niche sources, such as anonymous blogs; 2) *commerce*: business-related sources and online shops; 3) *government*: sources associated with local and national government agencies; 4) *healthcare*: sources associated with healthcare organizations, such as medical clinics and public health institutions; 5) *infotainment*: sources providing a mix of news and entertainment, such as HowStuffWorks; 6) *legacy media*: sources related to mainstream media organizations, such as CNN; 7) *non-government organization (NGO)*: sources attributed to mainstream non-government organizations; 8) *reference work*: online reference sources, such as Wikipedia; 9) *social media*: social media platforms, such as Reddit; 10) *think tank*: sources related to research institutions. The coding was conducted by the two authors, who consensus-coded disagreements between them.

Bibliography

- Battelle, J. (2005). The search: How Google and its rivals rewrote the rules of business and transformed our culture. Portfolio.
- Cardoso, B., & Magalhães, J. (2011). Google, Bing and a new perspective on ranking similarity. In

- Proceedings of the 20th ACM international conference on information and knowledge management (pp. 1933–1936). ACM Press.
- Cornett, L. (2007, August 17). Search & serendipity: Finding more when you know less. Search engine land. <https://searchengineland.com/search-serendipity-finding-more-when-you-know-less-11971>
- Cutrell, E., & Guan, Z. (2007). What are you looking for? An eye-tracking study of information usage in web search. In Proceedings of the SIGCHI conference on human factors in computing systems—CHI '07 (pp. 407–416). ACM Press.
- Foroohar, R. (2019). Don't be evil: The case against big tech. Penguin UK.
- González-Caro, C., & Marcos, M. C. (2011). Different users and intents: An eye-tracking analysis of web search. In Proceedings of WSDM 11 (pp. 9–12). ACM Press.
- Google Trends. (2020, April 21). covid-19, coronavirus. Google. <https://trends.google.com/trends/explore?q=covid-19,coronavirus>
- Hannak, A., Sapiezynski, P., Molavi Kakhki, A., Krishnamurthy, B., Lazer, D., Mislove, A., & Wilson, C. (2013). Measuring personalization of web search. In Proceedings of the 22nd international conference on World Wide Web (pp. 527–538). ACM Press.
- Harambam, J., Bountouridis, D., Makhortykh, M., & Van Hoboken, J. (2019). Designing for the better by taking users into account: A qualitative evaluation of user control mechanisms in (news) recommender systems. In Proceedings of the 13th ACM conference on recommender systems (pp. 69–77). ACM Press.
- He, C., Parra, D., & Verbert, K. (2016). Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications*, 56, 9–27.
- Helberger, N. (2011). Diversity by design. *Journal of Information Policy*, 1, 441–469.
- Jee, C. (2020, April 20). WhatsApp is limiting message forwarding to combat coronavirus misinformation. MIT technology review. <https://www.technologyreview.com/2020/04/07/998517/whatsapp-limits-message-forwarding-combat-coronavirus-misinformation/>
- Kemper, J., & Kolkman, D. (2019). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society*, 22(14), 2081–2096.
- Kliman-Silver, C., Hannak, A., Lazer, D., Wilson, C., & Mislove, A. (2015). Location, location, location: The impact of geolocation on web search personalization. In Proceedings of the 2015 internet measurement conference (pp. 121–127). ACM Press.
- Mittelstadt, B. (2016). Automation, algorithms, and politics: Auditing for transparency in content personalization systems. *International Journal of Communication*, 10: 4991–5002.
- Pasquale, F. (2015). *The black box society*. Harvard University Press.
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
- Puschmann, C. (2019). Beyond the bubble: Assessing the diversity of political search results. *Digital Journalism*, 7(6), 824–843.
- Robertson, R. E., Jiang, S., Joseph, K., Friedland, L., Lazer, D., & Wilson, C. (2018). Auditing partisan audience bias within Google Search. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW): 1–22.
- Schofield, J. (2019, December 12). Can DuckDuckGo replace Google search while offering better privacy? The Guardian. Retrieved from <https://www.theguardian.com/technology/askjack/2019/dec/12/duckduckgo-google-search-engine-privacy>
- Statista. (2020). Worldwide desktop market share of leading search engines from January 2010 to January 2020. <https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/>

- Webber, W., Moffat, A., & Zobel, J. (2010). A similarity measure for indefinite rankings. *ACM Transactions on Information Systems*, 28(4), 1–38.
- Weinberg, G. (2012). Privacy. DuckDuckGo <https://duckduckgo.com/privacy>

Funding

The research was conducted as part of the research project “Reciprocal relations between populist radical-right attitudes and political information behavior: A longitudinal study of attitude development in high-choice information environments,” funded by the Deutsche Forschungsgemeinschaft (project number 413190151) and Swiss National Science Foundation (project number 182630).

Competing interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethics

The data for the project were obtained from publicly available sources and thus were exempt from IRB review.

Copyright

This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

Data Availability

All materials needed to replicate this study are available via the Harvard Dataverse <https://doi.org/10.7910/DVN/IJKIN3>.

Acknowledgement

We would like to thank two anonymous reviewers and journal editors for their very helpful feedback on our manuscript, and student assistant of the Institute of Communication and Media Studies (University of Bern), Madleina Ganzeboom, for the valuable assistance during data collection.

Appendix

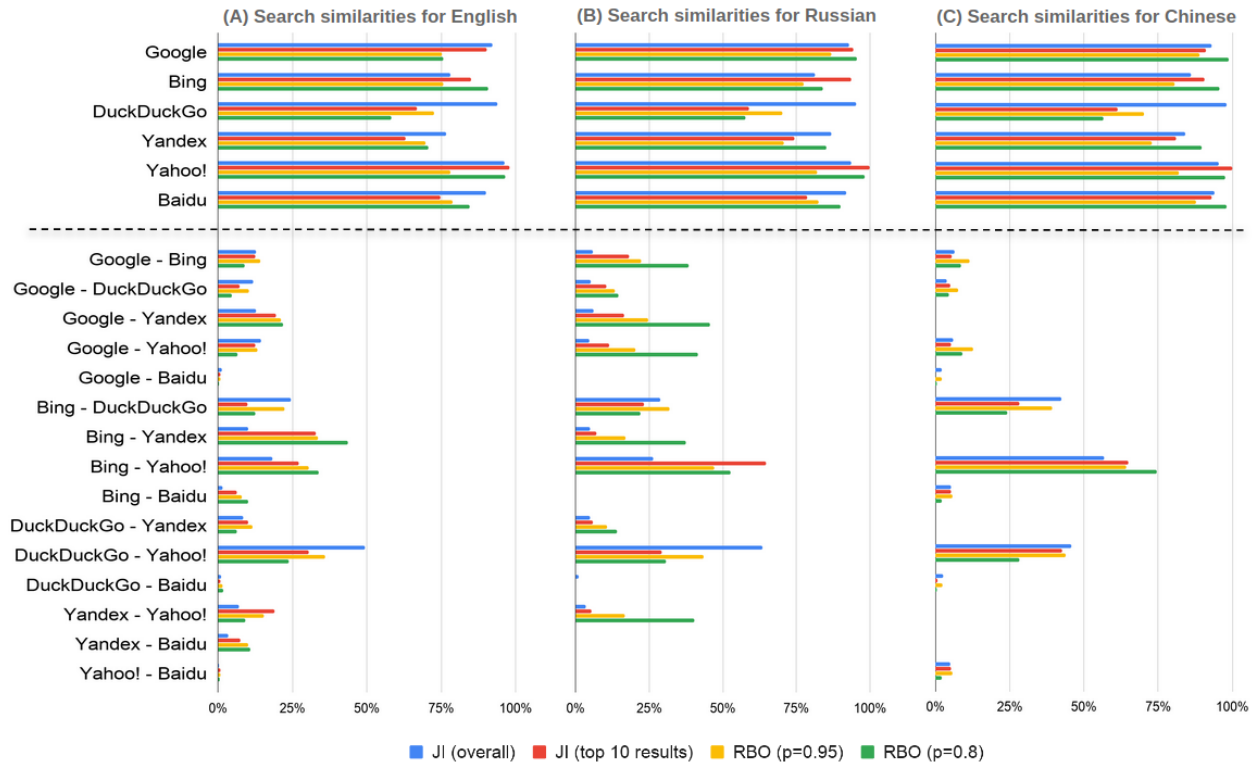


Figure 1A. Similarity metrics for the "US elections" query results. JI stands for the Jaccard index and RBO stands for the Ranked Biased Overlap. The x-axis shows the average percentage for each of the similarity metrics, and the y-axis shows the pairs of search engines that are compared.

Browser	Data type
Chrome	appcache, cache, cacheStorage, fileSystems, formData, history, indexedDB, pluginData, passwords, serviceWorkers, webSQL. For detailed information, see: https://developer.chrome.com/extensions/browsingData
Firefox	cache, formData, history, indexedDB, pluginData, passwords. For detailed information, see: https://developer.mozilla.org/en-US/d

Table 1. Types of data that were cleaned after each query round